



Foundations of Statistics & Probability
Module 01 Introduction to Statistics

Descriptive Statistics

Dr. Rajlaxmi Chouhan
Associate Professor, Department of Electrical Engineering
Head, Center for Education & Technology for Education
IIT Jodhpur



Step 1 Organizing the data

Example: Trends in streaming platform preferences among students

You surveyed **60 students** to find their favorite platform.

The options include popular platforms like **Netflix, Amazon Prime, Disney+ Hotstar, YouTube, Zee5, and SonyLIV**. The results are as follows:

Netflix: 12 votes

Amazon Prime: 10 votes

Disney+ Hotstar: 15 votes

YouTube: 12 votes

Zee5: 7 votes

SonyLIV: 4 votes

Step 1 Organizing the data

Netflix: 12 votes

Amazon Prime: 10 votes

Disney+ Hotstar: 15 votes

YouTube: 12 votes

Zee5: 7 votes

SonyLIV: 4 votes

Frequency Table



Streaming Platform	Frequency (Number of Votes)
Netflix	12
Amazon Prime	10
Disney+ Hotstar	15
YouTube	12
Zee5	07
SonyLIV	04

Step 2 Visualizing the data

Frequency Table

Streaming Platform	Frequency (Number of Votes)
Netflix	12
Amazon Prime	10
Disney+ Hotstar	15
YouTube	12
Zee5	07
SonyLIV	04



Line Graph

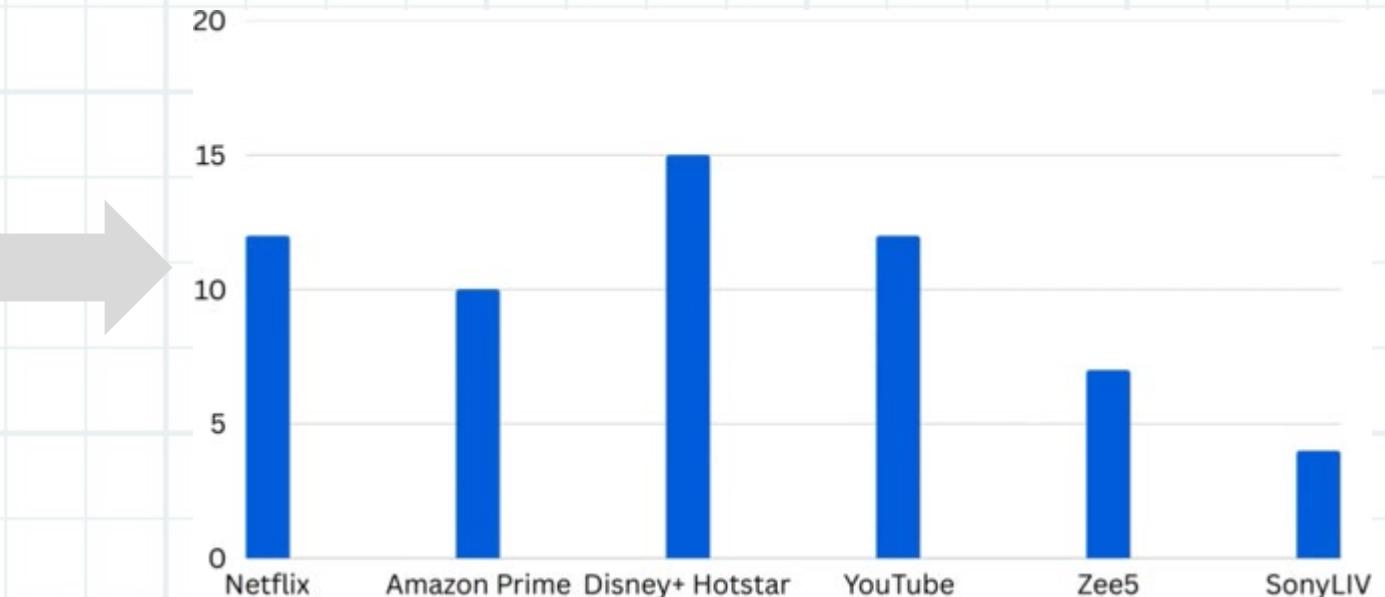


Step 2 Visualizing the data

Frequency Table

Streaming Platform	Frequency (Number of Votes)
Netflix	12
Amazon Prime	10
Disney+ Hotstar	15
YouTube	12
Zee5	07
SonyLIV	04

Bar Graph

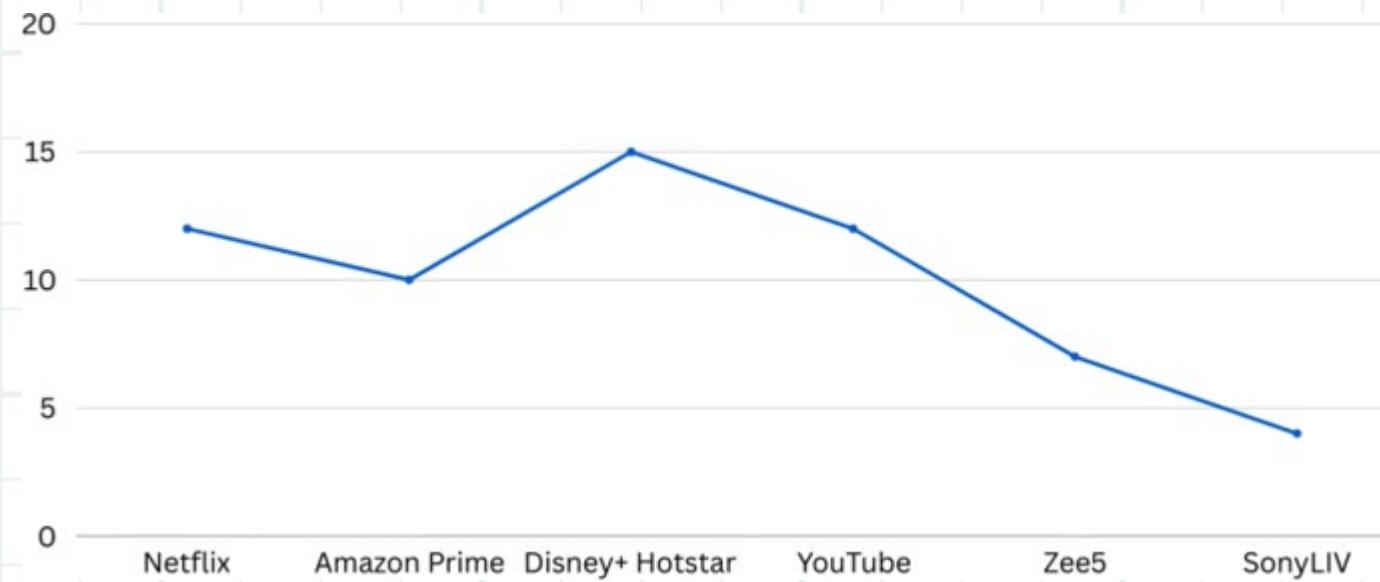


Step 2 Visualizing the data

Frequency Table

Streaming Platform	Frequency (Number of Votes)
Netflix	12
Amazon Prime	10
Disney+ Hotstar	15
YouTube	12
Zee5	07
SonyLIV	04

Frequency Polygon



Step 2 Visualizing the data

Consider a data set consisting of n values. If f is the frequency of a particular value, then the ratio f/n is called its *relative frequency*.

Frequency Table

Streaming Platform	Frequency (Number of Votes)
Netflix	12
Amazon Prime	10
Disney+ Hotstar	15
YouTube	12
Zee5	07
SonyLIV	04



Relative Frequency Table

Streaming Platform	Relative Frequency
Netflix	
Amazon Prime	
Disney+ Hotstar	
YouTube	
Zee5	
SonyLIV	

Step 2 Visualizing the data

Consider a data set consisting of n values. If f is the frequency of a particular value, then the ratio f/n is called its *relative frequency*.

Frequency Table

Streaming Platform	Frequency (Number of Votes)
Netflix	12
Amazon Prime	10
Disney+ Hotstar	15
YouTube	12
Zee5	07
SonyLIV	04



Relative Frequency Table

Streaming Platform	Relative Frequency
Netflix	12/60
Amazon Prime	10/60
Disney+ Hotstar	15/60
YouTube	12/60
Zee5	07/60
SonyLIV	04/60

Step 2 Visualizing the data

Consider a data set consisting of n values. If f is the frequency of a particular value, then the ratio f/n is called its *relative frequency*.

Frequency Table

Streaming Platform	Frequency (Number of Votes)
Netflix	12
Amazon Prime	10
Disney+ Hotstar	15
YouTube	12
Zee5	07
SonyLIV	04



Relative Frequency Table

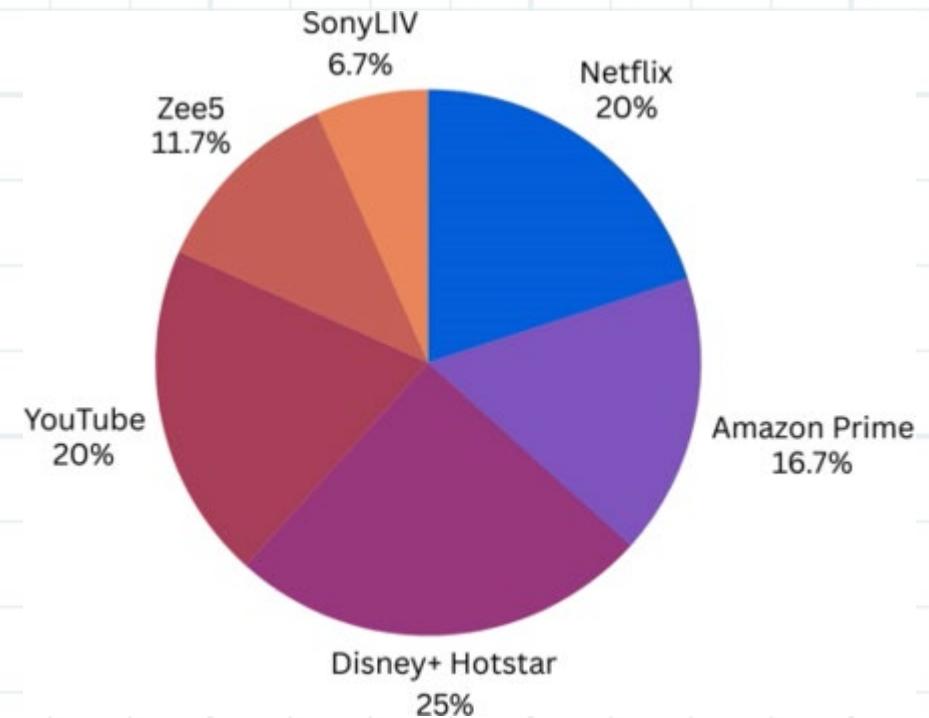
Streaming Platform	Relative Frequency
Netflix	20 %
Amazon Prime	16.67 %
Disney+ Hotstar	25 %
YouTube	20 %
Zee5	11.67 %
SonyLIV	6.67 %

Step 2 Visualizing the data

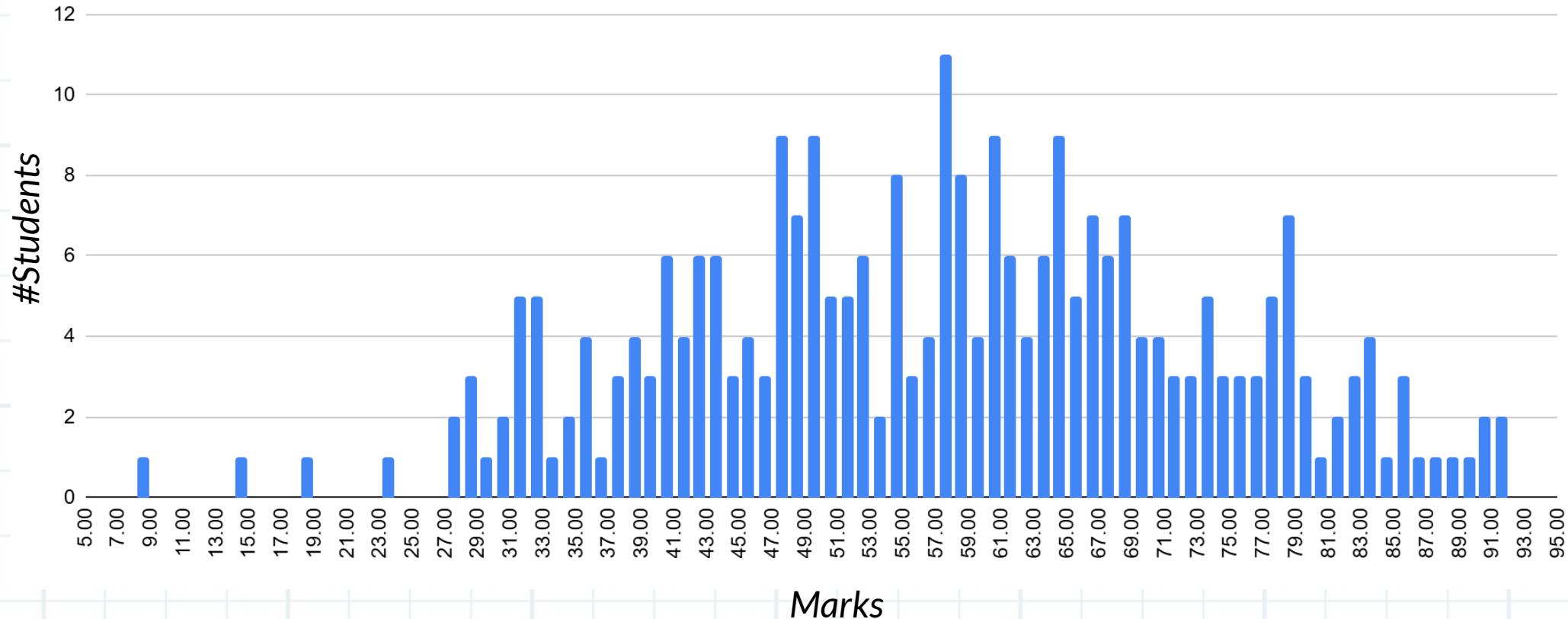
Relative Frequency Table

Streaming Platform	Relative Frequency
Netflix	20 %
Amazon Prime	16.67 %
Disney+ Hotstar	25 %
YouTube	20 %
Zee5	11.67 %
SonyLIV	6.67 %

Pie Chart

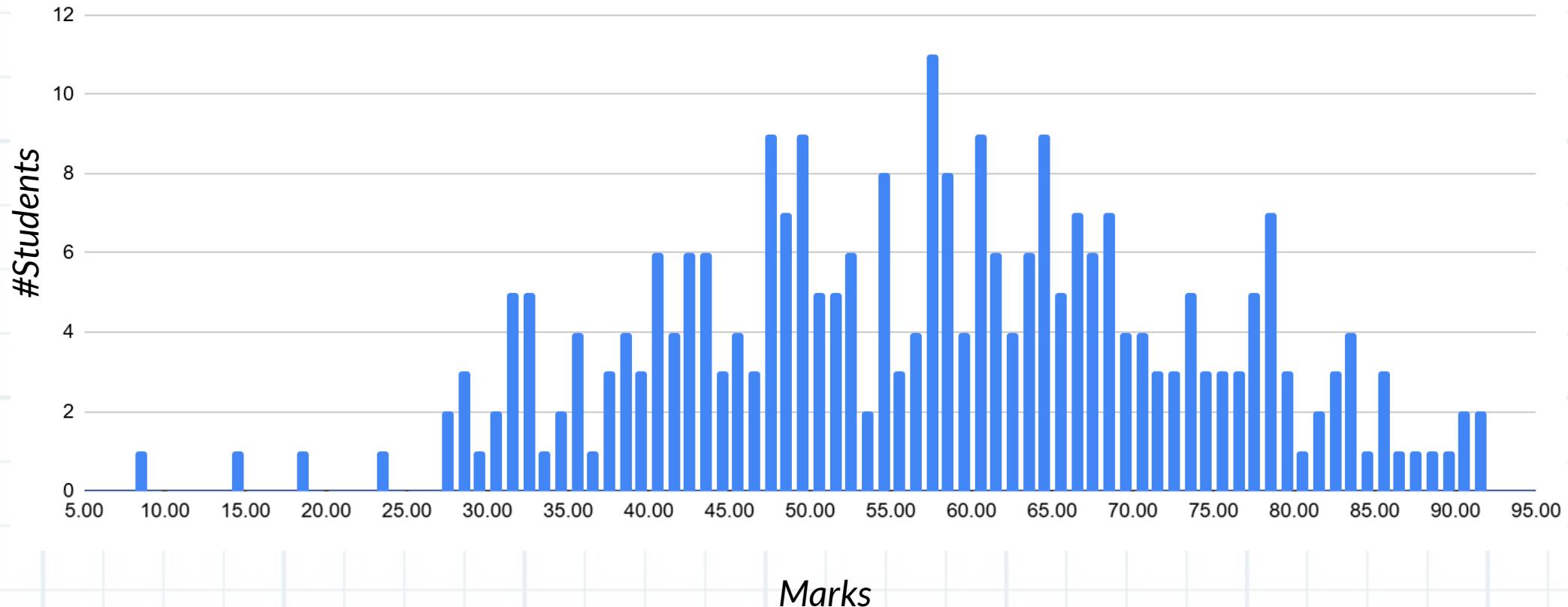


Grouped Data & Histograms

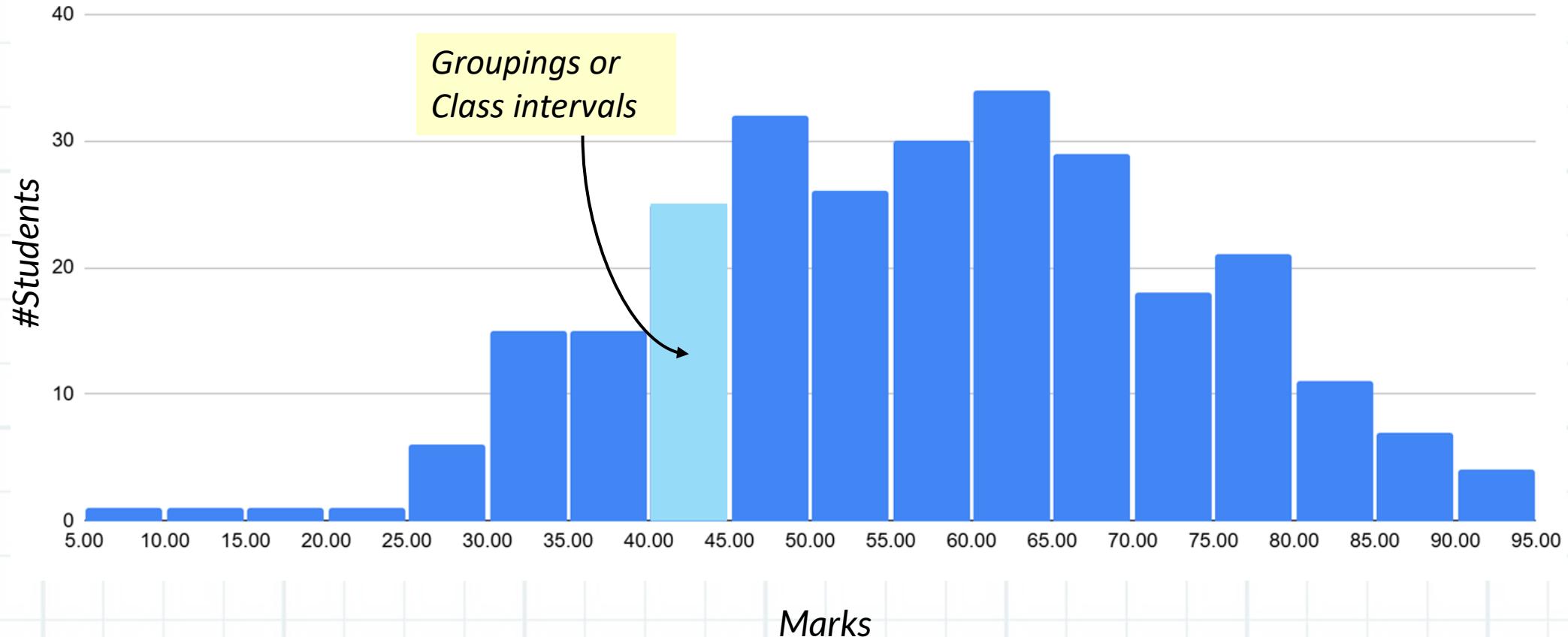




Grouped Data & Histograms

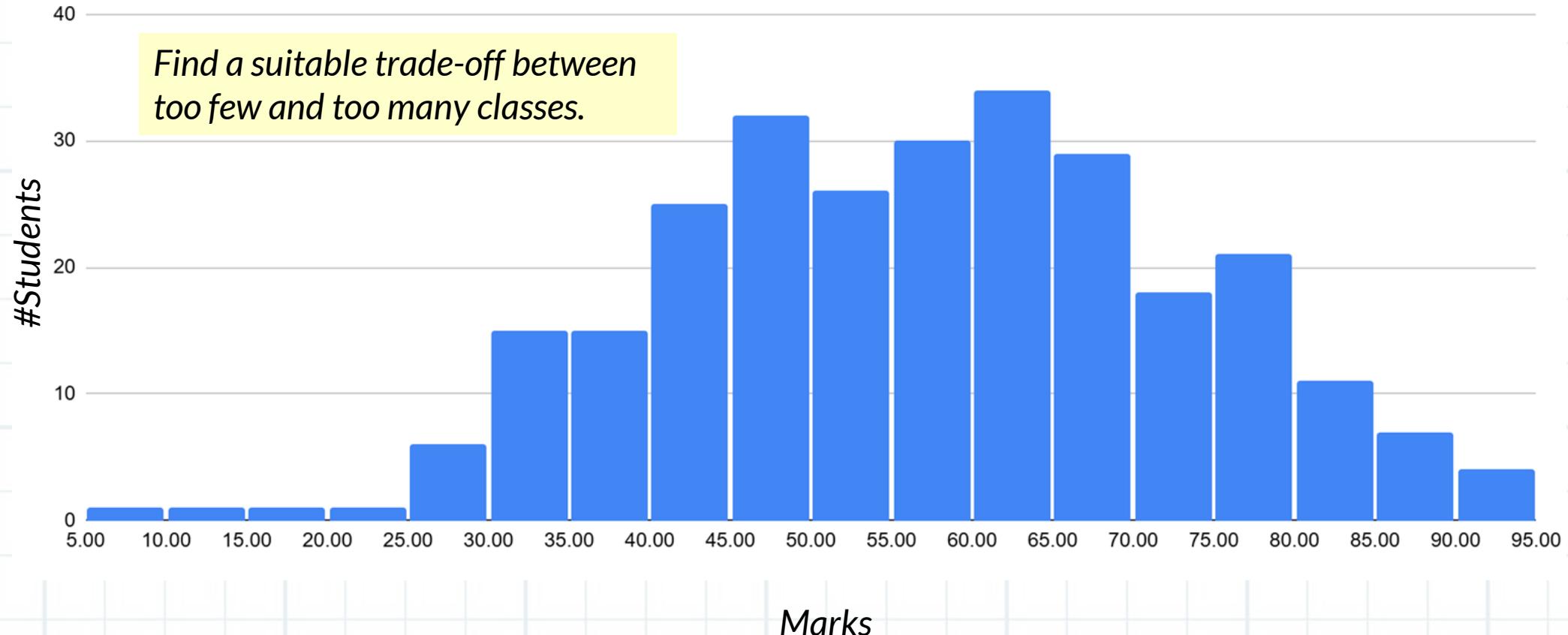


Histograms

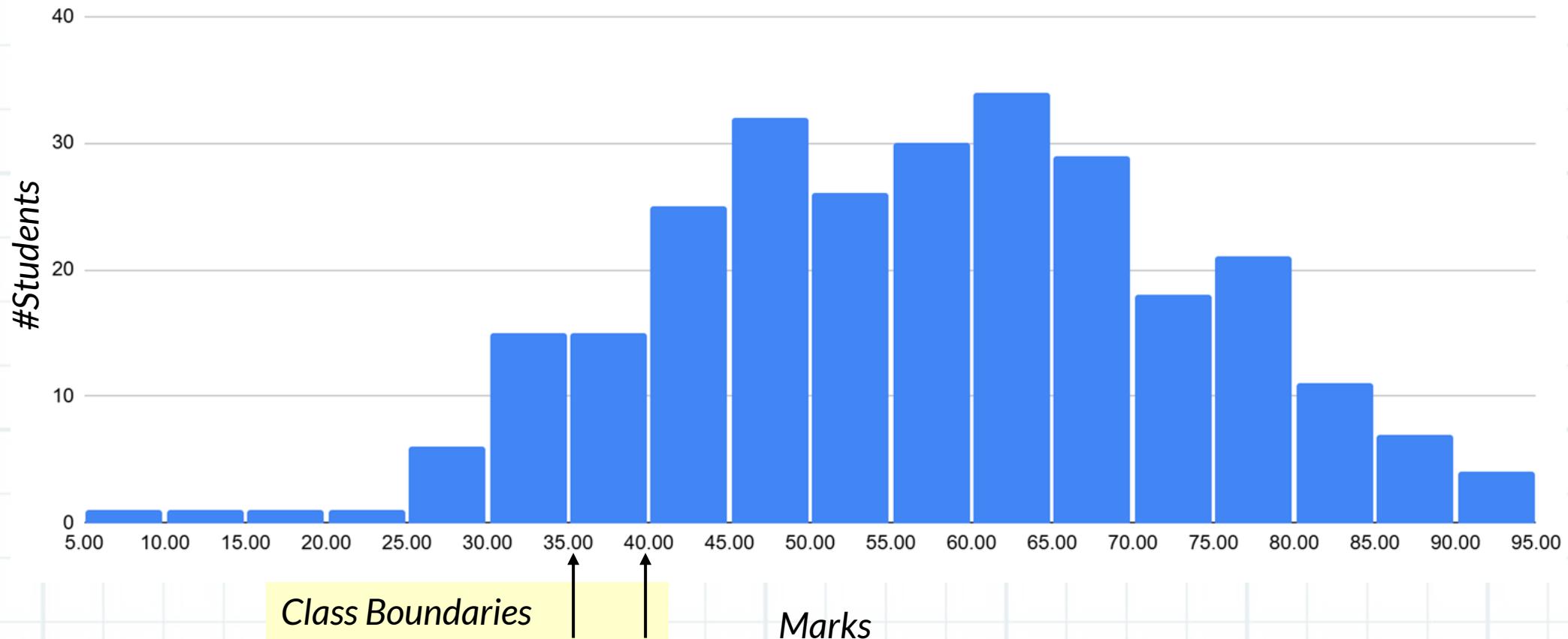




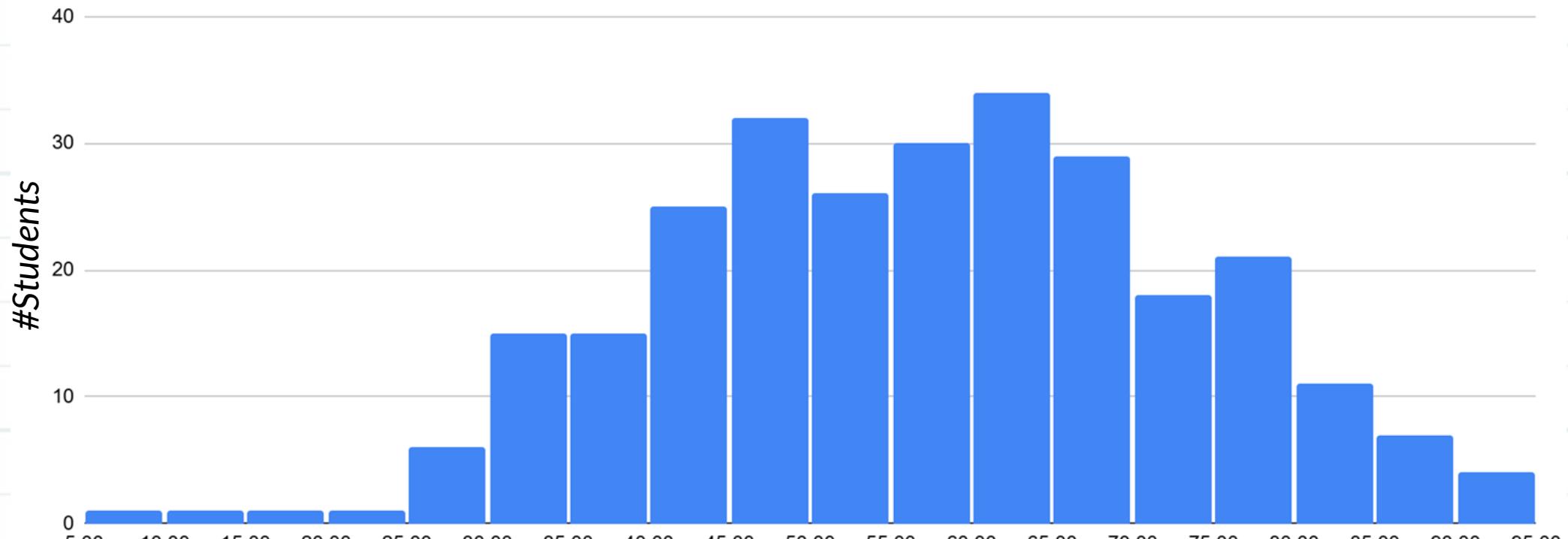
Histograms



Histograms



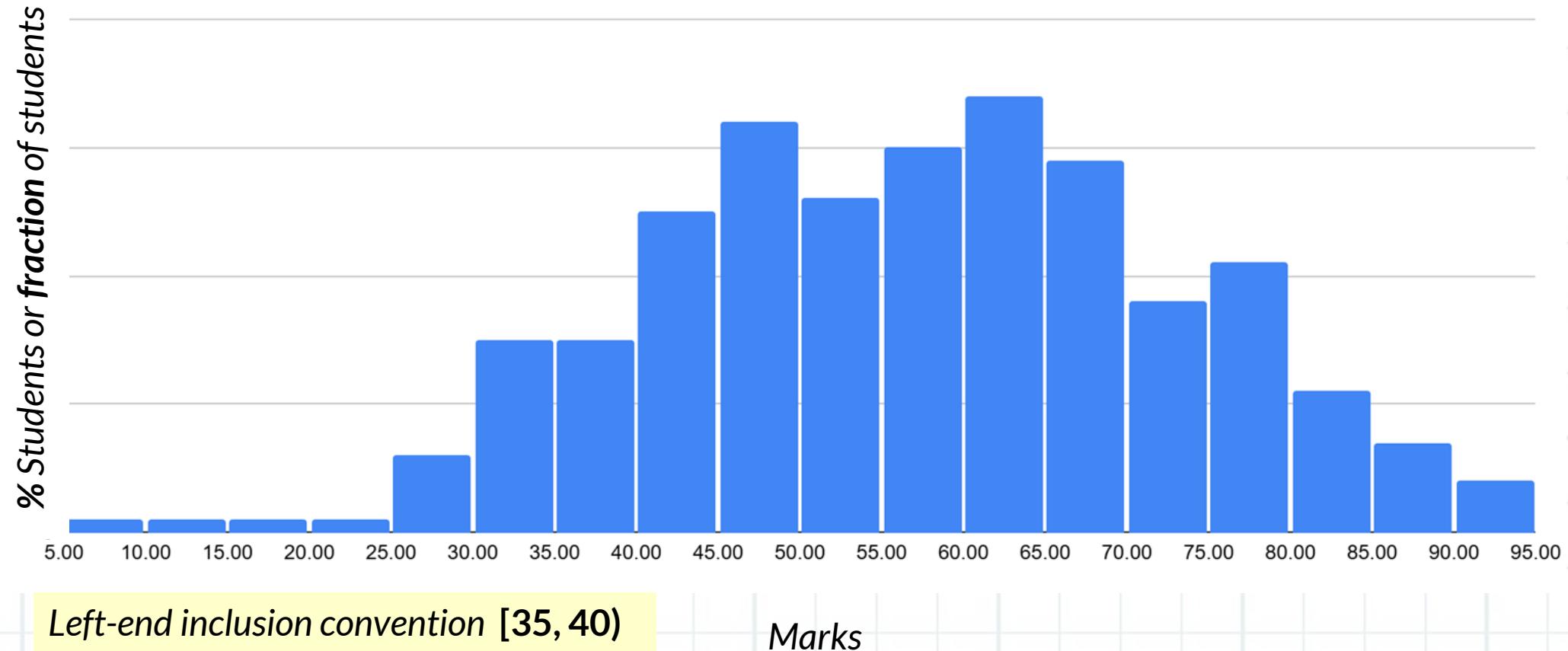
Histograms



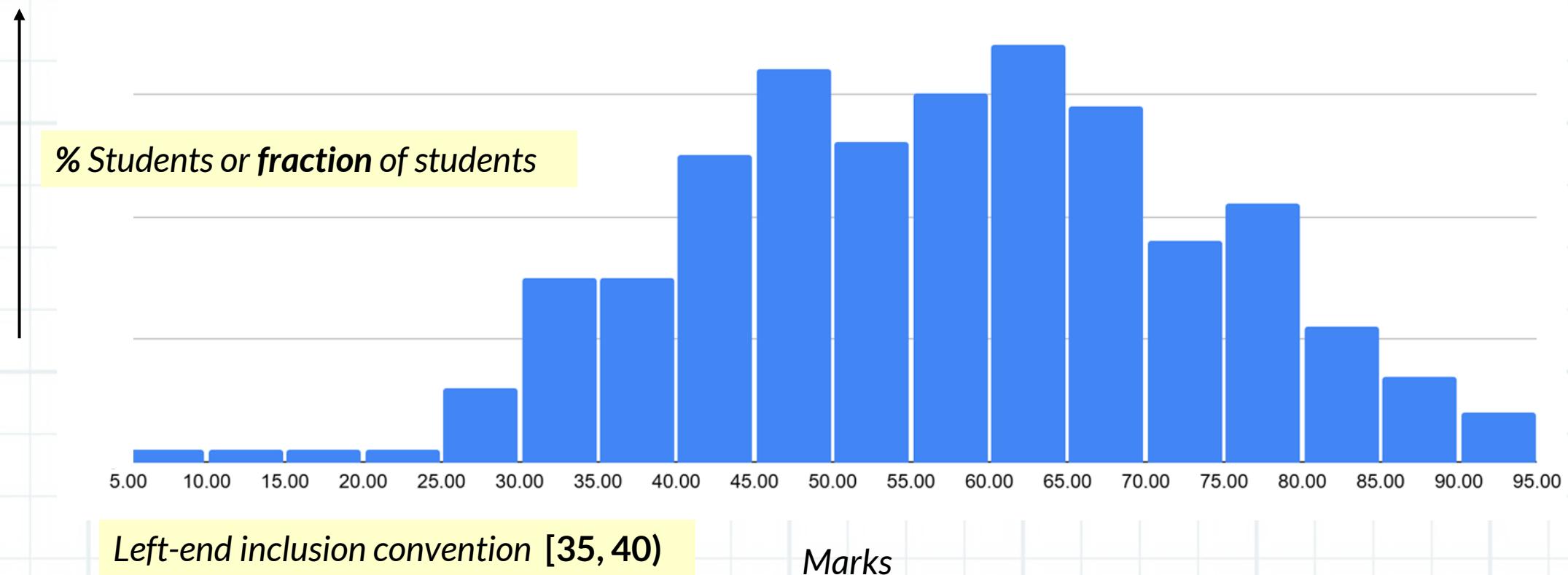
Left-end inclusion convention [35, 40)

Marks

Relative Frequency Histogram



Relative Frequency Histogram



Example

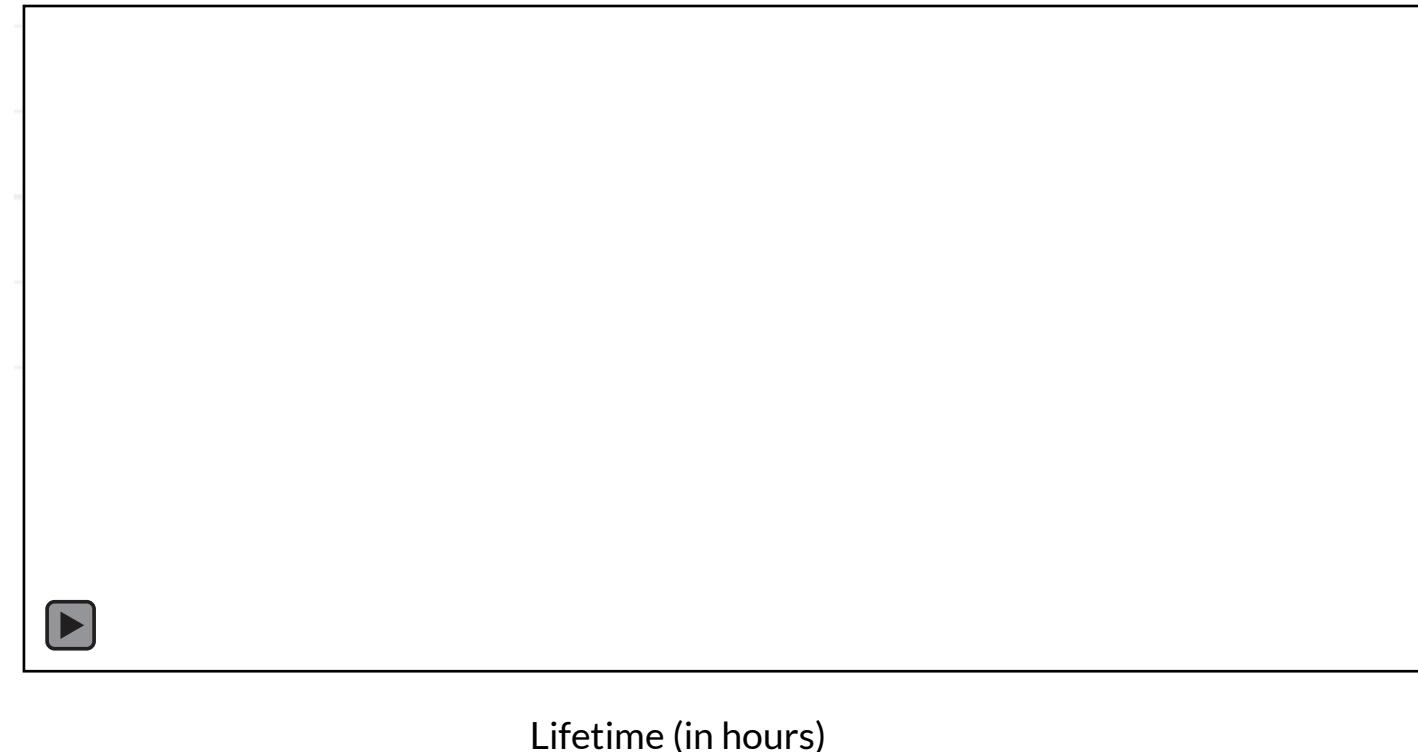
Life (in hours) of 200 Bulbs									
1067	919	1196	785	1126	936	918	1156	920	948
855	1092	1162	1170	929	950	905	972	1035	1045
1157	1195	1195	1340	1122	938	970	1237	956	1102
1022	978	832	1009	1157	1151	1009	765	958	902
923	1333	811	1217	1085	896	958	1311	1037	702
521	933	928	1153	946	858	1071	1069	830	1063
930	807	954	1063	1002	909	1077	1021	1062	1157
999	932	1035	944	1049	940	1122	1115	833	1320
901	1324	818	1250	1203	1078	890	1303	1011	1102
996	780	900	1106	704	621	854	1178	1138	951
1187	1067	1118	1037	958	760	1101	949	992	966
824	653	980	935	878	934	910	1058	730	980
844	814	1103	1000	788	1143	935	1069	1170	1067
1037	1151	863	990	1035	1112	931	970	932	904
1026	1147	883	867	990	1258	1192	922	1150	1091
1039	1083	1040	1289	699	1083	880	1029	658	912
1023	984	856	924	801	1122	1292	1116	880	1173
1134	932	938	1078	1180	1106	1184	954	824	529
998	996	1133	765	775	1105	1081	1171	705	1425
610	916	1001	895	709	860	1110	1149	972	1002

Class Interval	Frequency (Number of Items)
500–600	2
600–700	5
700–800	12
800–900	25
900–1000	58
1000–1100	41
1100–1200	43
1200–1300	7
1300–1400	6
1400–1500	1

Example

Class Interval	#Items
500–600	2
600–700	5
700–800	12
800–900	25
900–1000	58
1000–1100	41
1100–1200	43
1200–1300	7
1300–1400	6
1400–1500	1

Histogram



Cumulative Frequency Distribution

Class Interval	#Items (n)	Cumulative Frequency (c)
500–600	2	2
600–700	5	7
700–800	12	19
800–900	25	44
900–1000	58	102
1000–1100	41	143
1100–1200	43	186
1200–1300	7	193
1300–1400	6	199
1400–1500	1	200

Cumulative Frequency Distribution

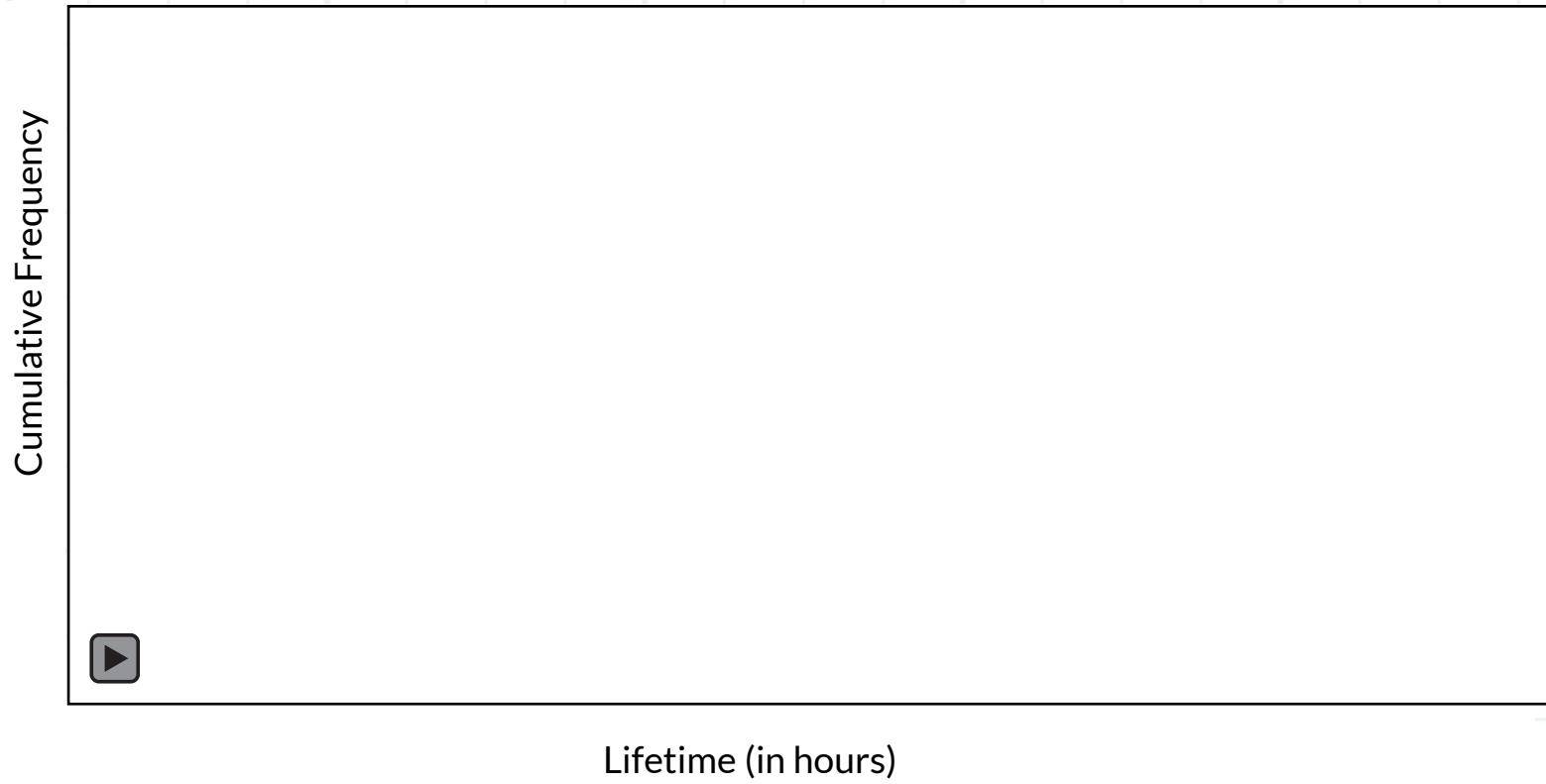
Class Interval	Cumulative Frequency (c)	Relative Cumulative Frequency (c / 200)
500–600	2	0.01
600–700	7	0.035
700–800	19	0.095
800–900	44	0.22
900–1000	102	0.51
1000–1100	143	0.715
1100–1200	186	0.93
1200–1300	193	0.965
1300–1400	199	0.995
1400–1500	200	1

Cumulative Frequency Distribution

Upper Boundary of Class	Cumulative Frequency (c)	Relative Cumulative Frequency (c / 200)
600	2	0.01
700	7	0.035
800	19	0.095
900	44	0.22
1000	102	0.51
1100	143	0.715
1200	186	0.93
1300	193	0.965
1400	199	0.995
1500	200	1

Cumulative Frequency Distribution

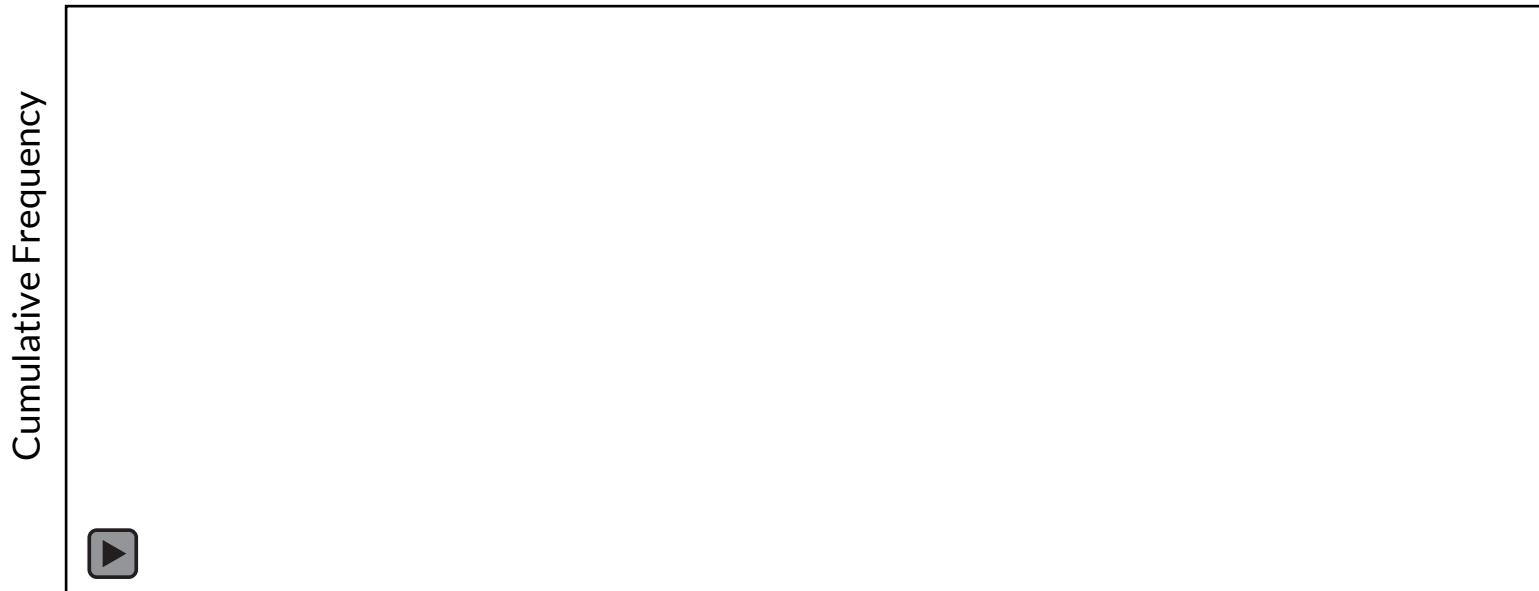
Upper Boundary of Class	Cumulative Frequency (c)
600	2
700	7
800	19
900	44
1000	102
1100	143
1200	186
1300	193
1400	199
1500	200



Cumulative Frequency Distribution

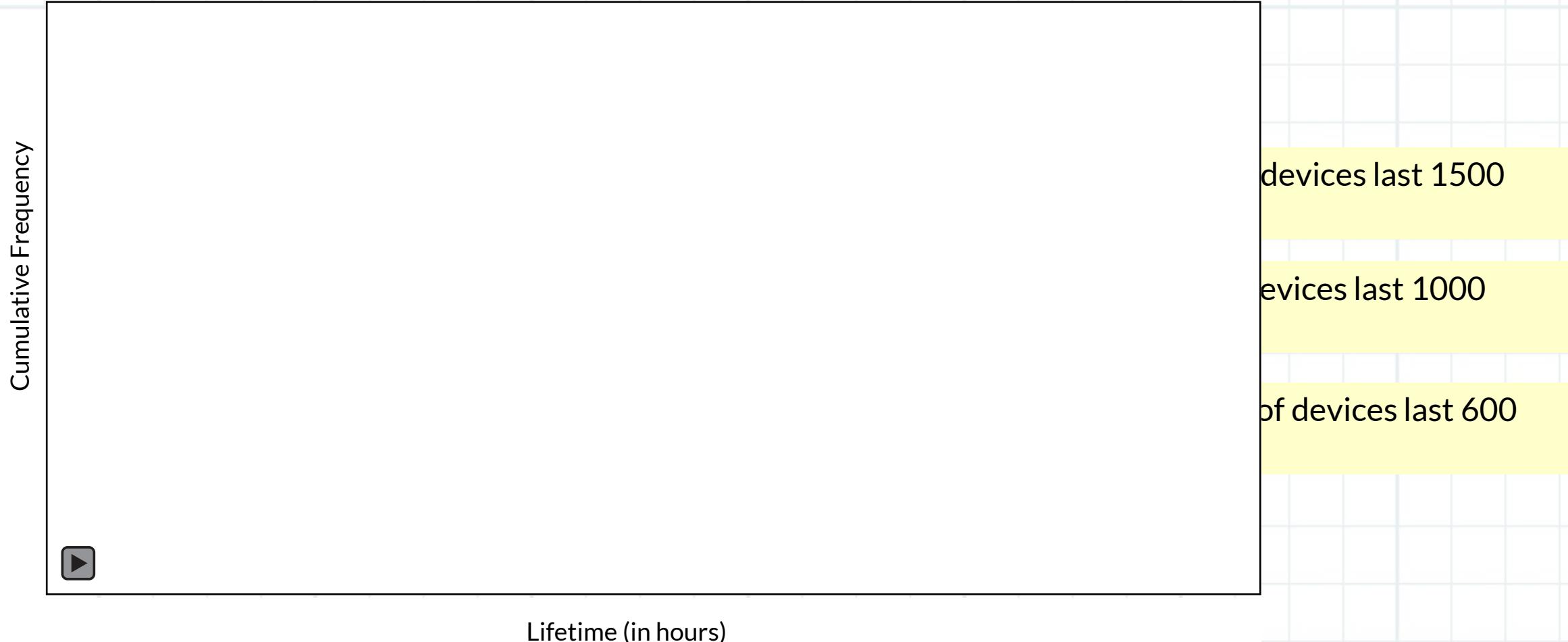
Upper Boundary of Class	Relative Cumulative Frequency
600	0.01
700	0.035
800	0.095
900	0.22
1000	0.51
1100	0.715
1200	0.93
1300	0.965
1400	0.995
1500	1

CDF or Ogive

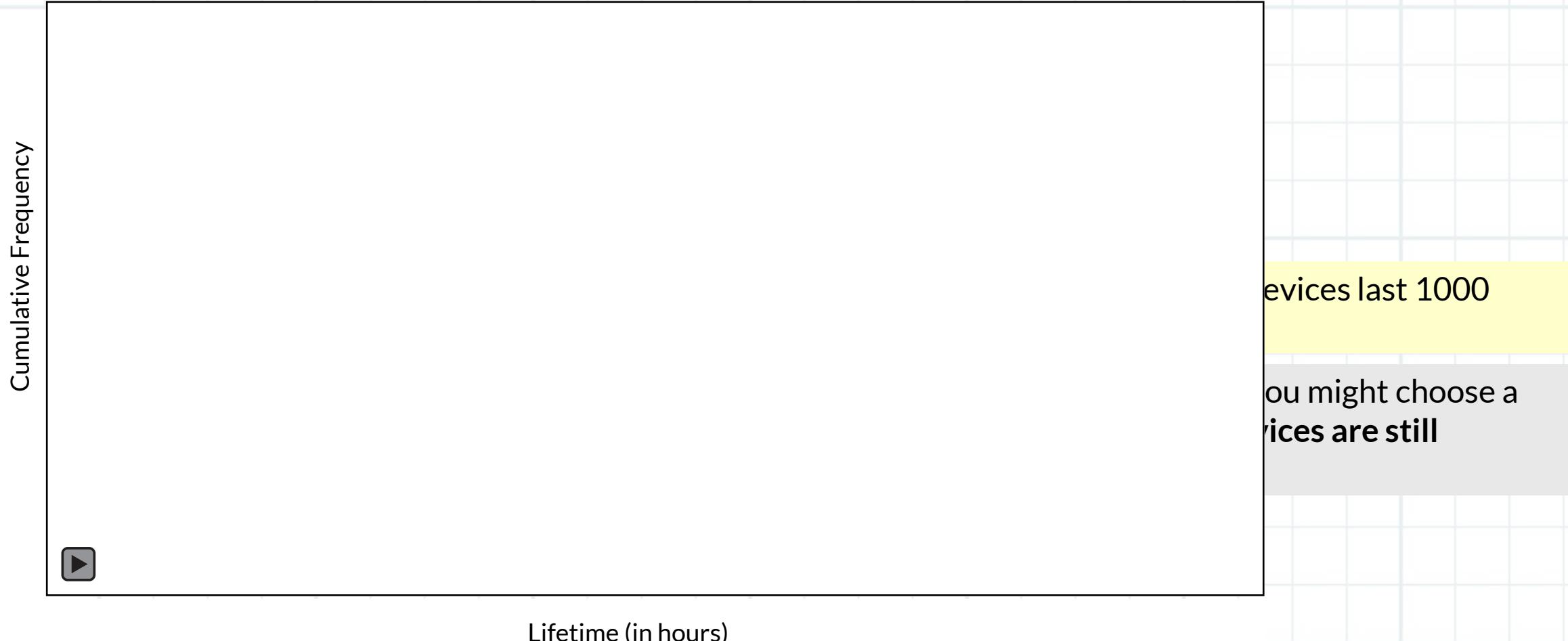


Lifetime (in hours)

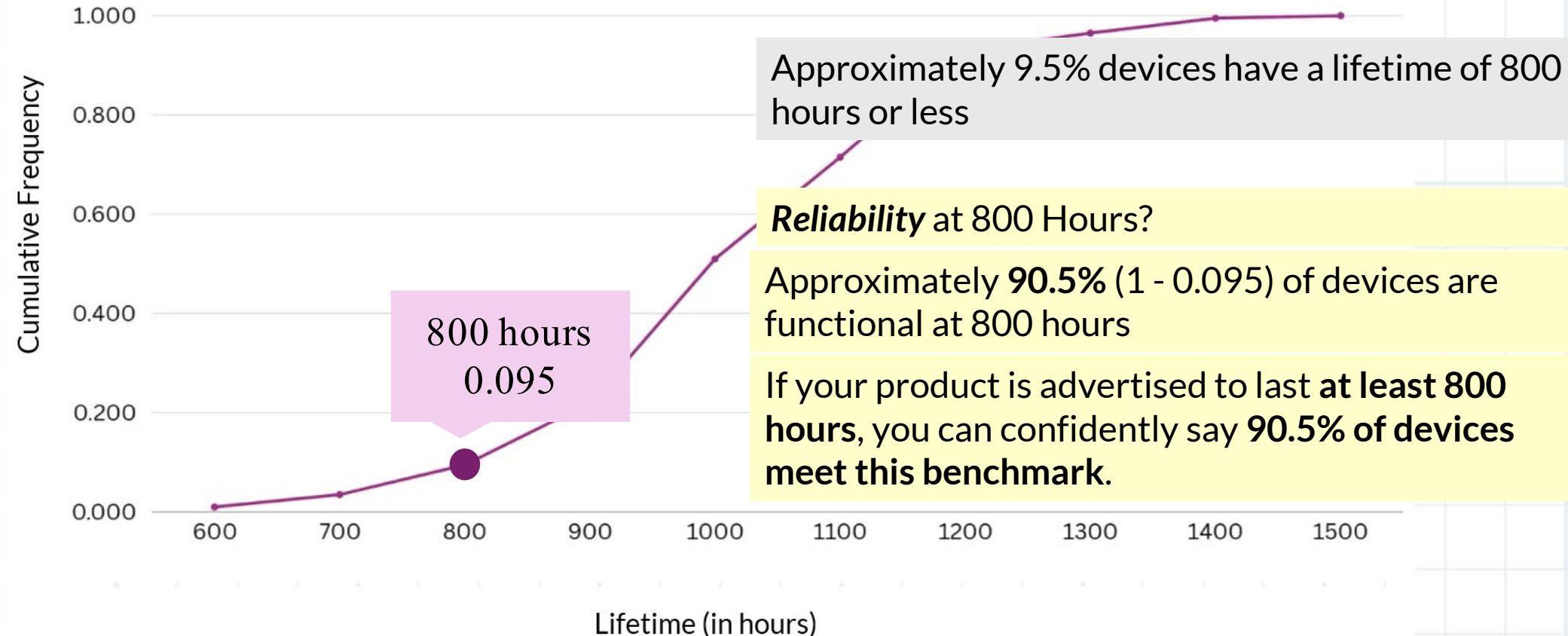
Key Observations



Key Observations



Key Observations



Recap



Organizing data
(frequency table, relative frequency table)

Visualizing data
(line graph, bar graph, frequency polygon, pie chart)

Grouped data
(Histogram and Cumulative Frequency Distribution)



Coming up next...

Descriptive Statistics Contd.

Summarizing data using mean, median and mode